Stopped Decision Processes with General Utility (Dynamic Decision Systems under Uncertain Environments)

AUTHOR(S):

Kadota, Yoshinobu; Kurano, Msami; Yasuda, Msasami

CITATION:

Kadota, Yoshinobu ...[et al]. Stopped Decision Processes with General Utility (Dynamic Decision Systems under Uncertain Environments). 数理解析研究所講究録 1998, 1048: 13-25

ISSUE DATE:

1998-05

URL:

http://hdl.handle.net/2433/62190

数理解析研究所講究録
1048巻 1998年 13-25

Stopped Decision Processes with General Utility

和歌山大教育　門田 良信 (Yoshinobu Kadota)
千葉大教育　蔵野 正美 (Masami Kurano)
千葉大理学　安田 正実 (Masami Yasuda)

abstract*

The general utility-treatment of stopped decision processes with a countable state space is considered.

Under reasonable conditions, the results of our previous paper[12] are extended to the general case, including the characterization of an optimal policy in case of the fixed stopping time.

For the case of the exponential utility functions, the optimal pair is sought concretely using the idea of the one-step look ahead (OLA) policy. Also, a numerical example is given.

1. Introduction

A combined model of the Markov decision process and the stopping problem, called stopped decision processes, has been considered by Furukawa and Iwamoto[6] in which the existence of an optimal pair of policy and stopping time associated with some optimality criterions is discussed for the reward system of the additive type.

Hordijk[8] has considered this model from a standpoint of potential theory.

Also, Furukawa[7] has reformulated the stopped decision model in the fashion of gambling theory and derived the optimality equation for the case of general recursive reward system using the successive approximation method.

In this paper the general utility-treatment of stopped decision processes with a countable state space is considered. Our previous paper[12] has already considered the optimization problem of the expected utility of the total discounted reward random variable accumulated until the stopping time and derived an optimality equation for the general utility case, by which an optimal pair of policy and stopping time has been characterized.

However, the concrete method of seeking an optimal pair is not discussed there.

* *Keywords* : Stopped Decision Processes, General Utility, Optimal Pairs, Optimality Equation, Exponential Utility.

The objective of this paper is to give the functional characterization from points of view of seeking an optimal pair. And we give further results concerning an optimality equation for our model, which is useful in seeking an optimal pair.

Also, for the case of the exponential utility function (cf. [5, 14]), the optimal pair is sought concretely using the idea of the one-step look ahead (OLA) policy(cf. [15]), giving a numerical example.

The optimality equations are described by the class of distribution functions of the present value, whose idea is appearing in Chung and Sobel[3], Sobel[16] and White[17].

In the remainder of this section, we shall formulate the problem to be examined. Also, an optimal pair of policy and stopping time is defined.

We consider standard Markov decision processes (MDPs), specified by

$$(S, \{A(i)\}_{i\in S}, q, r),$$

where $S = \{1, 2, \cdots\}$ denotes the set of the states of the processes, $A(i)$ is the set of actions available at each state $i \in S$, $q = (q_{ij}(a))$ is the matrix of transition probabilities satisfying that $\sum_{j\in S} q_{ij}(a) = 1$ for all $i \in S$ and $a \in A(i)$, and $r(i, a, j)$ is an immediate reward function defined on $\{(i, a, j) \mid i \in S, a \in A(i), j \in S\}$.

Throughout this paper we assume that (i) for each $i \in S$, $A(i)$ is a closed set of a compact metric space, (ii) for each $i, j \in S$, both $q_{ij}(\cdot)$ and $r(i, \cdot, j)$ are continuous on $A(i)$ and (iii) $r(\cdot, \cdot, \cdot)$ is uniformly bounded.

A sample space is the product space $\Omega = (S \times A)^\infty$ such that the projection X_t, Δ_t on the t-th factors S, A describe the state and the action of time t of the process ($t \geq 0$). A policy $\pi = (\pi_0, \pi_1, \cdots)$ is a sequence of conditional probabilities π_t such that $\pi_t(A(i_t) \mid i_0, a_0, \cdots, i_t) = 1$ for all histories $(i_0, a_0, \cdots, i_t) \in (S \times A)^t \times S$. The set of policies is denoted by Π.

Let $H_t = (X_0, \Delta_0, \cdots, \Delta_{t-1}, X_t)$ for $t \geq 0$. We assume for each $\pi = (\pi_0, \pi_1, \cdots) \in \Pi$,

$$P^\pi(X_{t+1} = j \mid H_{t-1}, \Delta_{t-1}, X_t = i, \Delta_t = a) = q_{ij}(a),$$

for all $t \geq 0$, $i, j \in S$, $a \in A(i)$. For any Borel measurable set X, $\mathcal{P}(X)$ denotes the set of all probability measures on X. Then, any initial probability measure $\nu \in \mathcal{P}(S)$ and policy $\pi \in \Pi$ determine the probability measure $P_\nu^\pi \in \mathcal{P}(\Omega)$ by a usual way.

The total present value until time t is defined by

$$\mathcal{B}(t) = \sum_{k=0}^{t} r(\, X_{k-1},\, \Delta_{k-1},\, X_k,\,) \quad (t \geq 0), \tag{1.1}$$

where X_{-1}, Δ_{-1} are fictitious variables and $r(\, X_{-1},\, \Delta_{-1},\, X_0,\,) = 0$.

Note that for each $\nu \in \mathcal{P}(S)$ and $\pi \in \Pi$, $\mathcal{B}(t)$ is a random variable from the probability space $(\, \Omega, P_\nu^\pi \,)$.

Let g be a non-decreasing continuous function on the real space $\mathcal{R}$.

Let $\nu \in \mathcal{P}(S)$ and $\pi \in \Pi$. Then, we call a random variable $\sigma : \Omega \longrightarrow \{0, 1, 2, \cdots\}$ a stopping time with respect to $(\, \nu,\, \pi\,)$ if the following conditions are satisfied:

(i) For each $t \geq 0$, $\{\sigma = t\} \in \mathcal{F}(H_t)$,

(ii) $P_\nu^\pi(\sigma < \infty) = 1$ and

(iii) $E_\nu^\pi[g^-(\mathcal{B}(\sigma))] < \infty$,

where $\mathcal{F}(H_t)$ is the σ-algebra induced by H_t and $g^-(x) = \max\{-g(x),\, 0\}$.

The set of such stopping times will be denoted by $\Sigma_{(\nu,\pi)}$.

For any $\nu \in \mathcal{P}(S)$, let

$$\mathcal{A}_\nu := \{(\pi, \sigma) \mid \sigma \in \Sigma_{(\nu,\pi)},\, \pi \in \Pi\}.$$

Then, interpreting a g as a utility function, our problem is to maximize the expected utility $E_\nu^\pi[g(\mathcal{B}(\sigma))]$ over all $(\pi,\, \sigma) \in \mathcal{A}_\nu$ for a fixed $\nu \in P(S)$.

The pair $(\pi^*, \sigma^*) \in \mathcal{A}_\nu$ is called $(\nu,\, g)$-optimal or simply optimal (suppressing $(\nu,\, g)$) if

$$E_\nu^{\pi^*}[g(\mathcal{B}(\sigma^*))] \geq E_\nu^\pi[g(\mathcal{B}(\sigma))] \quad \text{for all } (\pi, \sigma) \in \mathcal{A}_\nu. \tag{1.2}$$

In Section 2, we give the characterization of an optimal policy in the case that the stopping time is fixed, whose results are applied to obtain an optimal pair in the sequel.

In Section 3, we extend the results obtained in [12, 13] for the discount case to the general case. The proofs are nearly analogous to those in [12, 13], so that the most part of the proofs will be omitted.

The exponential utility case is treated in Section 4, where the optimal pair is sought by using the idea of the one-step look ahead (OLA) stopping time.

2. MDPs with the stopping region

For a subset K of S, let

$$\sigma_K := \text{ the first time } t \geq 0 \text{ such that } X_t \in K.$$

Henceforth, we assume that σ_K is a stopping time with respect to any $(\nu, \pi) \in \mathcal{P}(S) \times \Pi$.

We say that $\pi^* \in \Pi$ is (ν, g)-optimal with respect to σ_K if

$$E_\nu^{\pi^*}[g(\mathcal{B}(\sigma_K))] \geq E_\nu^{\pi}[g(\mathcal{B}(\sigma_K))] \quad \text{for all } \pi \in \Pi.$$

When π^* is (ν, g)-optimal for all $\nu \in \mathcal{P}(S)$, π^* is simply called g-optimal w.r.t. σ_K.

In order to analize the above problem, it is convenient to rewrite $E_\nu^{\pi}[g(\mathcal{B}(\sigma_K))]$ by using the distribution function of $\mathcal{B}(\sigma_K)$ corresponding to P_ν^π. Suppressing K in the notation, let for $\nu \in \mathcal{P}(S)$ and $\pi \in \Pi$,

$$F_\nu^\pi(z) := P_\nu^\pi(\, \mathcal{B}(\sigma_K) \leq z\,) \quad \text{and}$$
$$\Phi(\nu) := \{F_\nu^\pi(\cdot) \mid \pi \in \Pi\}.$$

Then, it is obvious that $E_\nu^{\pi}[g(\mathcal{B}(\sigma_K))] = \int g(z) F_\nu^\pi(dz)$.

For any $\pi \in \Pi$, the map $v_\pi : \mathcal{R} \times \mathcal{P}(S) \longrightarrow \mathcal{R}$ will be defined by

$$v_\pi(d, \nu) := \int g(d+z) F_\nu^\pi(dz).$$

We note that $v_\pi(0, \nu) = E_\nu^{\pi}[g(\mathcal{B}(\sigma_K))]$ and $v_\pi(d, i) = g(d)$ if $i \in K$, where $\nu \in \mathcal{P}(S)$ is simply denoted by i when it is degenerate at $\{i\}$.

The value function for our model can be denoted by

$$v(d, \nu) = \sup_{\pi \in \Pi} v_\pi(d, \nu), \tag{2.1}$$

which is depending on the present fortune $d \in \mathcal{R}$ and state distribution $\nu \in \mathcal{P}(S)$.

In the following lemma, it is shown that the supremum in (2.1) can be attainable.

Lemma 2.1. *For any* $\nu \in \mathcal{P}(S)$, $v(d, \nu) = \max_{F \in \Phi(\nu)} \int g(d+z) F(dz)$.
Also, for each $\nu \in \mathcal{P}(S)$, *there exists* (ν, g)*-optimal policy w.r.t.* σ_K.

Proof. For each $\nu \in \mathcal{P}(S)$, the set $\{\, P_\nu^\pi(\cdot) \in \mathcal{P}(\Omega) \mid \pi \in \Pi\,\}$ is known to be compact in the weak topology(c.f. Borker[1]). Since the map $\mathcal{B}(\sigma_K) : \Omega \longrightarrow \mathcal{R}$ is continuous, $\Phi(\nu)$ is

weak-compact. Thus, from the assumption of the continuity of $g_d(z) = g(d+z)$, it follows that

$$v(d, \nu) = \sup_{F\in\Phi(\nu)} \int g_d(z)F(dz) = \int g_d(z)F^*(dz) \quad \text{for some } F^* \in \Phi(\nu).$$

This proves the first part of the results.

For $F^* \in \Phi(\nu)$ with $v(0, \nu) = \int g(z)F^*(dz)$, the policy π^* coresponding to F^* is clearly (ν, g)-optimal w.r.t. σ_K, as required. □

Lemma 2.2. *For each $t \geq 0$, $d \in \mathcal{R}$ and $\pi \in \Pi$,*

$$\begin{aligned} &E_\nu^\pi[g(d+\mathcal{B}(\sigma_K)) \mid \sigma_K > t] \\ &\leq E_\nu^\pi\Big[\max_{a\in A(X_t)} \sum_{j\in S} q_{X_t,j}(a)v(d+\mathcal{B}(t)+r(X_t, a, j), j) \mid \sigma_K > t\Big]. \end{aligned} \tag{2.2}$$

Proof. For simplicity, denote E_ν^π by E. For any $\omega = (i_0, a_0, i_1, a_1, \cdots) \in \Omega$, let $\theta_t(\omega) = (i_t, a_t, i_{t+1}, \cdots)$ be a shift operator for $t \geq 1$.

The Markov property of the transition yields that

$$\begin{aligned} &E[g(d+\mathcal{B}(\sigma_K)) \mid \sigma_K > t] \\ &= E[E[g(d+\mathcal{B}(t+1)+\mathcal{B}(\sigma_K-(t+1))(\theta_{t+1}(\omega))) \mid H_{t+1}] \mid \sigma_K > t] \\ &\leq E[v(d+\mathcal{B}(t)+r(X_t, \Delta_t, X_{t+1}), X_{t+1}) \mid \sigma_K > t] \\ &\leq \{ \text{ the right-hand of the inequality (2.2) } \}, \end{aligned}$$

which completes the proof. □

For any $d \in \mathcal{R}$ and $i \notin K$, let

$$A(d, i) := \arg\max_{a\in A(i)} \sum_{j\in S} q_{ij}(a)v(d+r(i, a, j), j).$$

The value function $v(d, i)$ is shown to satisfy the optimality equation in the fillowing theorem.

Theorem 2.1. *The value function v satisfies the following equation.*

$$v(d, i) = \begin{cases} \max_{a\in A(i)} \sum_{j\in S} q_{ij}(a)v(d+r(i, a, j), j) & \text{for } i \notin K, \\ g(d) & \text{for } i \in K. \end{cases} \tag{2.3}$$

Proof. Let $d \in \mathcal{R}$. For any $f \in F$ such that $f(i) \in A(d, i)$ for $i \notin K$ and $f(i) = a(\text{arbitrary}) \in A(i)$ for $i \in K$, let $\pi^{(j)}$ be a policy corresponding to $F_j^* \in \Phi(j)$ satisfying

$$v(d+r(i, f(i), j), j) = \int g(d+r(i, f(i), j)+z)F_j^*(dz),$$

whose existence is guaranteed by Lemma 2.1.

Let π be the policy that chooses the action Δ_0 at time 0 according to f and use policy $\pi^{(j)}$ from time 1 when $X_1 = j$. Then, clearly it holds

$$\begin{aligned} E_i^{\pi}[g(d+\mathcal{B}(\sigma_K))] &= \sum_{j\in S} q_{ij}(f(i))v(d+r(i,\, f(i),\, j),\, j) \\ &= \max_{a\in A(i)} \sum_{j\in S} q_{ij}(a)v(d+r(i,\, a,\, j),\, j). \end{aligned}$$

Together with (2.2), the above derives (2.3), as required. □

In order to discuss the uniqueness of solutions of (2.3), we need the following assumption.

Assumption A. $E_\nu^{\pi}[|\, g(d+\mathcal{B}(\sigma_K))\,|\,] < \infty$ for any $\nu \in \mathcal{P}(S)$, $\pi \in \Pi$ and $d \in \mathcal{R}$.

Theorem 2.2. *Suppose that Assumption A holds. Then,*

(i) it follows that

$$\lim_{t\to\infty} E_\nu^{\pi}[v(d+\mathcal{B}(t),\, X_t)\mathbf{1}_{\{\sigma_K>t\}}] = 0 \tag{2.4}$$

for any $\nu \in \mathcal{P}(S)$, $\pi \in \Pi$ and $d \in \mathcal{R}$, where $\mathbf{1}_A$ is an indicator function of a set A.

(ii) *The map $v : \mathcal{R}\times S \longrightarrow \mathcal{R}$ satisfying (2.4) is uniquely determined by (2.3).*

Proof. Let $\pi \in \Pi$. Let $\pi\{X_t\} \in \Pi$ be such that $v(d+\mathcal{B}(t),\, X_t) = v_{\pi\{X_t\}}(d+\mathcal{B}(t),\, X_t)$. Note that $\pi\{X_t\}$ is depending on $d+\mathcal{B}(t)$ and X_t and its existence is guaranteed by Lemma 2.1.

We denote by $\pi^{(t)} \in \Pi$ the policy that uses π until time t and uses $\pi\{X_t\}$ from time t. Then,

$$\begin{aligned} E_\nu^{\pi^{(t)}}[g(d+\mathcal{B}(\sigma_K))] &= E_\nu^{\pi}[g(d+\mathcal{B}(\sigma_K))\mathbf{1}_{\{\sigma_K\le t\}}] \\ &\quad + E_\nu^{\pi}[v(d+\mathcal{B}(t),\, X_t)\mathbf{1}_{\{\sigma_K>t\}}]. \end{aligned} \tag{2.5}$$

Under Assumption A,

$$\lim_{t\to\infty} E_\nu^{\pi^{(t)}}[g(d+\mathcal{B}(\sigma_K))] = E_\nu^{\pi}[g(d+\mathcal{B}(\sigma_K))].$$

So as $t \to \infty$ in (2.5), we get (i).

The proof of (ii) is not particularly difficult, but tedious. So, we omit it. □

The following results can be proved as similarly as that of Theorem 3.3 in [10].

Theorem 2.3. *Let* $\pi^* = (\pi_0^*, \pi_1^*, \cdots)$ *be any policy satisfying that for all* $t \geq 0$

$$\pi_t^*(A(\mathcal{B}(t), X_t) \mid H_t) = 1 \text{ on } \{\sigma_K > t\}.$$

Then π^* *is g-optimal w.r.t.* σ_K.

3. Optimal pairs

In this section we derive the optimality equation for the stopped decision model, by which an optimal pair is characterized. Throughout this section, we assume the following Conditions 1 and 2 are satisfied.

Condition 1. The utility function g is differential and for any compact subset D of $\mathcal{R}$, there exists a constant L_D such that

$$| g'(x) | \leq L_D \quad \text{for all } x \in D.$$

Condition 2. $E_\nu^\pi[\sup_{t\geq 0} g^+(\mathcal{B}(t))] < \infty$ for all $\nu \in \mathcal{P}(S)$ and $\pi \in \Pi$.

For simplicity of the notations, let

$$\widehat{\Phi}(\nu) := \{F_\nu^{(\pi,\sigma)} \mid (\pi, \sigma) \in \mathcal{A}_\nu\},$$

where $F_\nu^{(\pi,\sigma)}(x) = P_\nu^\pi(\mathcal{B}(\sigma) \leq x)$ for $(\pi, \sigma) \in \mathcal{A}_\nu$.

In order to describe an optimality equation in the sequel, define:

$$U\{g\}(d, i, a, j) = \sup_{F \in \widehat{\Phi}(j)} \int g(d + r(i, a, j) + z) F(dz) \tag{3.1}$$

and

$$U\{g\}(d, i) = \max_{a \in A(i)} \sum_{j \in S} q_{ij}(a) U\{g\}(d, i, a, j) \tag{3.2}$$

for each $d \in \mathcal{R}$, $i, j \in S$ and $a \in A(i)$.

It is easily proved under Condition 1 that the maximum in (3.2) is attainable.

For $\nu \in \mathcal{P}(S)$ and $n \geq 1$, let

$$\mathcal{A}_\nu^n := \{(\pi, n \vee \sigma) \mid (\pi, \sigma) \in \mathcal{A}_\nu\}.$$

where $a \vee b = \max\{a, b\}$ for $a, b \in \mathcal{R}$.

Define a conditional maximum by

$$\gamma_n^\nu := \text{esssup}_{(\pi,\sigma)\in\mathcal{A}_\nu^n} E_\nu^\pi[g(\mathcal{B}(\sigma)) \mid \mathcal{F}_n] \quad (n \geq 0),$$

where $\mathcal{F}_n = \mathcal{F}(H_n)$.

Henceforth, for simplicity we write esssup by sup and suppress ν in γ_n^ν if not specified otherwise. The recursive relation concerning $\{\gamma_n\}$ is described in the following, whose proof is given in [13].

Lemma 3.1. ([12, 13]) *For each $n \geq 0$, it holds*

(i) $\gamma_n = \max\{\ g(\mathcal{B}(n)),\ \sup_{\pi\in\Pi} E_\nu^\pi[\gamma_{n+1} \mid \mathcal{F}_n]\}$

(ii) $\sup_{\pi\in\Pi} E_\nu^\pi[\gamma_{n+1}|\mathcal{F}_n] = U\{g\}(\mathcal{B}(n),\ X_n)$.

In order to obtain an optimal pair, it is convenient to introduce the following notations:

$$\boldsymbol{R} := \{(d,\ i) \in \mathcal{R} \times S \mid\ g(d) \geq U\{g\}(d,\ i)\ \} \quad \text{and}$$

$$A^*(d,\ i) := \arg\max_{a\in A(i)} \sum_{j\in S} q_{ij}(a) U\{g\}(d,\ i,\ a,\ j)$$

for each $d \in \mathcal{R}$ and $i \in S$.

Let σ^* is the first time $t \geq 0$ such that

$$(\mathcal{B}_t,\ X_t) \in \boldsymbol{R} \tag{3.3}$$

and $\pi^* = (\pi_0^*,\ \pi_1^*, \cdots)$ be any policy satisfying

$$P_\nu^{\pi^*}(\ \Delta_t \in A^*(\mathcal{B}(t),\ X_t)) = 1 \quad \text{for all} \quad t \geq 0. \tag{3.4}$$

The following lemma is given in [12, 13].

Lemma 3.2. ([12, 13]) *Let $\sigma^*(n) = \min\{\sigma^*,\ n\}$. Then, $\{\ \gamma_{\sigma^*(n)},\ \mathcal{F}_n,\ n \geq 0\ \}$ is a martingale.*

Here, we can state the main theorem.

Theorem 3.1.

(i) *If $P_\nu^{\pi^*}(\sigma^* < \infty) = 1$, then the pair (π^*, σ^*) is g-optimal.*

(ii) *If $g(\mathcal{B}(n)) \longrightarrow -\infty$ (as $n \to \infty$) $P_\nu^{\pi^*}$-a.s. then $P_\nu^{\pi^*}(\sigma^* < \infty) = 1$.*

Proof. From Lemma 3.2, $E_\nu^{\pi^*}[\gamma_0] = E_\nu^{\pi^*}[\gamma_{\sigma^*(n)}]$ for all $n \geq 1$. Now, as $n \to \infty$ in the above, we get

$$
\begin{aligned}
&E_\nu^{\pi^*}[\gamma_{\sigma^*}\mathbf{1}_{\{\sigma^*<\infty\}}] + E_\nu^{\pi^*}[\underline{\lim}_{n\to\infty}\gamma_{\sigma^*(n)}\mathbf{1}_{\{\sigma^*=\infty\}}] \\
&\leq E_\nu^{\pi^*}[\gamma_0] \leq E_\nu^{\pi^*}[\gamma_{\sigma^*}\mathbf{1}_{\{\sigma^*<\infty\}}] + E_\nu^{\pi^*}[\overline{\lim}_{n\to\infty}\gamma_{\sigma^*(n)}\mathbf{1}_{\{\sigma^*=\infty\}}].
\end{aligned} \tag{3.5}
$$

If $P_\nu^{\pi^*}(\sigma^* < \infty) = 1$, $E_\nu^{\pi^*}[\gamma_0] = E_\nu^{\pi^*}[\gamma_{\sigma^*}]$.

On the other hand, by the definition of σ^*, $\gamma_{\sigma^*} = g(\mathcal{B}(\sigma^*))$, which implies $E_\nu^{\pi^*}[\gamma_0] = E_\nu^{\pi^*}[g(\mathcal{B}(\sigma^*))]$.

Since $E_\nu^{\pi}[g(\mathcal{B}(\sigma))] \leq E_\nu^{\pi}[\gamma_0] = E_\nu^{\pi^*}[\gamma_0]$, it holds $E_\nu^{\pi}[g(\mathcal{B}(\sigma))] \leq E_\nu^{\pi^*}[g(\mathcal{B}(\sigma^*))]$ for all $(\pi, \sigma) \in \mathcal{A}_\nu$. This shows that the pair (π^*, σ^*) is g-optimal.

(ii) follows obviously from the right inequality of (3.5). □

4. Exponential utility functions

In this section, we consider the case of the exponential utility function

$$g_\lambda(x) = \operatorname{sign}(-\lambda)\exp(-\lambda x) \tag{4.1}$$

for a non-zero constant λ and try to give the concrete characterization of the optimal pair by the idea of the one-step look ahead (OLA) stopping time. For the OLA-stopping time, refer to Ross[15] and Kadota et.al[11].

Let

$$\eta_\lambda(i, a) := \sum_{j\in S} q_{ij}(a)\exp(-\lambda r(i, a, j)).$$

We need the following Conditions.

Condition A. For any $\lambda > 0$, $\eta_\lambda(i, a)$ is non-decreasing in $i \in S$ for each $a \in A$, and for $\lambda < 0$, $\eta_\lambda(i, a)$ is non-increasing in $i \in S$ for each $a \in A$.

Condition B. For each $a \in A$, $q_{ij}(a) = 0$, if $i > j$ and $q_{ii}(a) < 1$.

We note that Condition B is satisfied for Markov deteriorating system.

Let

$$K_\lambda := \{(d, i) \in \mathcal{R} \times S \mid v_\lambda(d, i) - g_\lambda(d) \leq 0\},$$

where

$$v_\lambda(d, i) = \max_{a\in A(i)} \sum_{j\in S} q_{ij}(a) g_\lambda(d + r(i, a, j)).$$

Then, the K_λ is characterized by the following.

Lemma 4.1. *Under Condition A, for each λ ($\lambda \neq 0$), there exists an integer $i_\lambda \in S$ such that $K_\lambda = \mathcal{R} \times \{i \in S \mid i \geq i_\lambda\}$.*

Proof. We observe that $v_\lambda(d, i) - g_\lambda(d) = e^{-\lambda d}(1 - \min_{a \in A(i)} \eta_\lambda(i, a))$ if $\lambda > 0$, $= e^{-\lambda d}(\max_{a \in A(i)} \eta_\lambda(i, a) - 1)$ if $\lambda < 0$.

So that, if $\lambda > 0$, $v_\lambda(d, i) - g_\lambda(d) \leq 0$ means $\min_{a \in A} \eta_\lambda(i, a) \geq 1$.

Observing that $\min_{a \in A} \eta_\lambda(i, a)$ is non-decreasing in $i \in S$ from Condition A, there exists an i_λ such that $v_\lambda(d, i) - g_\lambda(d) \leq 0$ if and only if $i \geq i_\lambda$. Similarly the case of $\lambda < 0$ is proved, as required. □

Lemma 4.2. *Under Condition A and B, the following holds:*

(i) *If $i \geq i_\lambda$, then*

$$\sum_{j \in S} q_{ij}(a) U\{g_\lambda\}(d, i, a, j) = \sum_{j \in S} q_{ij}(a) g_\lambda(d + r(i, a, j)) \tag{4.2}$$

for all $a \in A(i)$ and $d \in \mathcal{R}$.

(ii) *If $i < i_\lambda$, then $g_\lambda(d) < U\{g_\lambda\}(d, i)$ for all $d \in \mathcal{R}$.*

Proof. Let $i \geq i_\lambda$, $a \in A(i)$, $d \in \mathcal{R}$ and $j \in S$ with $q_{ij}(a) > 0$. Then, for any $F \in \widehat{\Phi}(j)$, from Condition B and Lemma 4.1 we see that $\{ g_\lambda(d + r(i, a, j) + \mathcal{B}(n)), \mathcal{F}_n, n = 0, 1, 2, \cdots \}$ is a suppermartingale with respect to P_j^π, where (π, σ) is a pair corresponding to F and $\mathcal{F}_n = \mathcal{F}(H_n)$.

Applying Theorem 2.2 of Chow, Robbins and Siegmund[2], we get

$$\begin{aligned} E_j^\pi[g_\lambda(d + r(i, a, j) + \mathcal{B}(\sigma))] &= \int g_\lambda(d + r(i, a, j) + z) F(dz) \\ &\leq g_\lambda(d + r(i, a, j)), \end{aligned}$$

which means $U\{g_\lambda\}(d, i, a, j) \leq g_\lambda(d + r(i, a, j))$. Thus (4.2) follows.

For (ii), let $i < i_\lambda$. Then, for any $d \in \mathcal{R}$, since $(d, i) \notin K_\lambda$ there exists $a_1 \in A(i)$ such that $g(d) < \sum_{j \in S} q_{ij}(a_1) g_\lambda(d + r(i, a_1, j))$. Thus, by the definition, clearly $g(d) < U\{g_\lambda\}(d, i)$, as required. □

From Lemma 4.2, we find that the optimal stopping time σ_λ^* defined by (3.3) in Section 3 becomes

$$\sigma_\lambda^* = \text{ the first time } t \geq 0 \text{ with } X_t \in K_\lambda,$$

which is σ_{K_λ} with the stopping region K_λ and discussed in Section 2. Thus, to seek the optimal policy π^*, we can apply the results in Section 2.

Let

$$v_\lambda^*(i) = \mathrm{Opt}_{\pi\in\Pi} E_i^\pi[e^{-\lambda\mathcal{B}(\sigma_\lambda^*)}]$$

where "Opt" means "Maximum" if $\lambda < 0$ and "Minimum" if $\lambda > 0$. Then, by Theorem 2.1 we have :

(4.3)

$$v_\lambda^*(i) = \begin{cases} \mathrm{Opt}_{a\in A(i)}\Big[\sum\limits_{i\le j<i_\lambda} q_{ij}(a)e^{-\lambda r(i,a,j)}\cdot v_\lambda^*(j) + \sum\limits_{j\ge i_\lambda} q_{ij}(a)e^{-\lambda r(i,a,j)}\Big], & \text{for } i < i_\lambda, \\ 1, & \text{for } i \ge i_\lambda. \end{cases}$$

Let, for i $(1 \le i < i_\lambda)$,

$$A^*(i) = \{a \in A(i) \mid a \text{ realizes the opt on the right-hand side of (4.3)}\}$$

Then, the optimal pair $(\pi_\lambda^*, \sigma_\lambda^*)$ under exponential utility is given in the following theorem.

Theorem 4.1. *Let σ_λ^* = the first t such that $X_t \ge i_\lambda$ and $\pi_\lambda^* = (\pi_0^*, \pi_1^*, \cdots)$ be such that $\pi_t^*\{A^*(X_t) \mid H_t\} = 1$ for $1 \le X_t < i_\lambda$.*

Then, the pair $(\pi_\lambda^*, \sigma_\lambda^*)$ is g-optimal.

Proof. We can check that $\boldsymbol{R}$ and $A^*(d, i)$ in Section 3 is equal to K_λ and $A^*(i)$ respectively. Thus, from Theorem 3.1, Theorem 4.1 follows. □

Here we give a numerical example to illustrate the theoretical results.

Let $S = \{1, 2, 3, \cdots\}$, $A = [1,\ 2]$ and

$$q_{ij}(a) = \begin{cases} \dfrac{(\frac{a}{i})^{j-i}}{(j-i)!}e^{-\frac{a}{i}}, & j \ge i \\ 0 & j < i, \end{cases}$$

for $i, j \in S$ and $a \in A$.

For an inspection cost $c > 0$, let $r(i, a, j) = \frac{a}{i} - c$ $\quad (i, j \in S,\ a \in A)$. Then, $\eta_\lambda(i, a) = e^{-\lambda(\frac{a}{i}-c)}$, which satisfies Condition A. Simple calculations yield the integer i_λ in Lemma 4.1 is given as $i_\lambda = \lceil\frac{2}{c}\rceil$, which is independent of λ, where $\lceil x\rceil$ is the smallest integer greater than or equal to x.

Also, by (4.3) we find $A^*(i) = \{2\}$, so that the optimal policy $\pi_\lambda^* = 2^\infty$.

As another example, let $r(i,\ a,\ j) = \frac{a}{i} \mid j - i \mid -c$. Then,

$$\eta_\lambda(i,\ a) = e^{\lambda c} \exp\{\frac{a}{i}(\exp(-\frac{\lambda a}{i}) - 1)\},$$

which satisfies Condition A. The numerical value of each integer i_λ is given in Table 1. Observing Table 1, we know that a risk-averse decision maker ($\lambda > 0$) has a tendency to stop earlier than a risk-seeking one ($\lambda < 0$).

λ		-2.5	-2	-1.5	-1	-0.5		0.5	1	1.5	2	2.5
i_λ	c=0.1	8	8	8	7	7		7	6	6	6	6
	c=0.01	22	21	21	21	21		20	20	20	19	19

Table 1: The value of i_λ for $c = 0.1$ and $0.01 (\lambda \neq 0)$.

References

[1] V. S. Borkar, *Topics in Controlled Markov Chains*, Longman Scientific Technical, 1991.

[2] Y. S. Chow, H. Robbins and D. Siegmund, *The Theory of Optimal Stopping : Great Expectations*, Houghton Mifflin Company, 1971.

[3] K. J. Chung and M. J. Sobel, Discounted MDP's: Distribution functions and exponential utility maximization, *SIAM J. Control Optim.* **25**(1987), 49-62.

[4] E.V. Denardo and U.G. Rothblum, Optimal stopping, exponential utility and linear programming, *Math. Prog.* **16**, 1979, pp.228–244.

[5] P.C. Fishburn, *Utility Theory for Decision Making*, John Wiley & Sons, New York, 1970.

[6] N. Furukawa and S. Iwamoto, Stopped decision processes on complete separable metric spaces, *J. Math. Anal. Appl.*, **31**, 1970, 615–658.

[7] N.Furukawa, Functional equations and Markov potential theory in stopped decision processes, *Mem. Fac. Sci, Kyushu Univ. Ser. A* **29**, 1975, 329–347.

[8] A. Hordijk, Dynamic Programming and Potential Theory, *Math. Centre Tracts No.* **51**, Mathematisch Centrum, Amsterdam, 1974.

[9] R.S. Howard and J.E. Matheson, Risk-sensitive Markov decision processes, *Management Sci.* **8**, 1972, pp.356–369.

[10] Y. Kadota, M. Kurano and M. Yasuda, Discounted Markov decision processes with general utility functions, *Proceedings of APORS' 94*, World Scientific, 330–337, 1995.

[11] Y. Kadota, M. Kurano and M. Yasuda, Utility-Optimal Stopping in a denumerable Markov chain, *Bull. Inform. Cybernet.* **28**(1), 1995, pp.15–21.

[12] Y. Kadota, M. Kurano and M. Yasuda, On the general utility of discounted Markov decision processes. To appear in *International Transactions in Operational Reseach*, **4**, 1998.

[13] Y. Kadota, Studies on risk-sensitive Markov decision processes with a countable state space, Ph. D. Dissertation, Chiba University, 1998.

[14] J.W. Pratt, Risk aversion in the small and in the large, *Econometrica* **32**, 1964, 122–136.

[15] S.M. Ross, *Applied Probability Models with Optimization Applications*, Holden-Day, 1970.

[16] M.J. Sobel, The variance of discounted Markov decision processes, *J. Appl. Prob.* **19**, 1982, 794–802.

[17] D.J. White, Minimizing a threshold probability in discounted Markov decision processes, *J. Math. Anal. Appl.* **173** (1993) 634–646.